Human Genomics

RESEARCH

Open Access

High clinical utility of long-read sequencing for precise diagnosis of congenital adrenal hyperplasia in 322 probands

Yunpeng Wang[1,2,3†], Gaohui Zhu[1,2†], Danhua Li[4†], Yu Pan[1,2], Rong Li[1,2], Ting Zhou[1,2], Aiping Mao[4], Libao Chen[4], Jing Zhu[3,5*] and Min Zhu[1,2*]

Abstract

Background The molecular genetic diagnosis of congenital adrenal hyperplasia (CAH) is very challenging due to the high homology between the *CYP21A2* gene and its pseudogene *CYP21A1P*.

Methodology This study aims to assess the clinical efficacy of targeted long-read sequencing (T-LRS) by comparing it with a control method based on the combined assay (NGS, Multiplex ligation-dependent probe amplification and Sanger sequencing) and to introduce T-LRS as a first-tier diagnostic test for suspected CAH patients to improve the precise diagnosis of CAH.

Results A large cohort of 562 participants including 322 probands and 240 family members was enrolled for the perspective (96 probands) and prospective study (226 probands). The comparison analysis of T-LRS and control method have been performed. In the perspective study, 96 probands were identified using both the control method and T-LRS. Concordant results were detected in 85.42% (82/96) of probands. T-LRS performed more precise diagnosis in 14.58% (14/96) of probands. Among these, a novel 4141 kb deletion involving *CYP21A2* and *TNXB* was established. A new diagnosis was improved by T-LRS. The duplications were also precisely identified to clarify the misdiagnosis by MLPA. In the prospective study, Variants were identified not only in *CYP21A2* but also in *HSD3B2* and *CYP11B1* in 226 probands. Expand to 322 probands, the actual frequency of duplication haplotype (1.55%) could be calculated due to the accurate genotyping. Moreover, 75.47% of alleles with SNVs/indels, 22.20% of alleles with deletion chimeras.

Conclusion T-LRS has higher resolution and reduced cost than control method with accurate diagnosis. The clinical utility of L-LRS could help to provide precision therapy to CAH patients, advance the life-long management of this complex disease and promote our understanding of CAH.

Keywords Congenital adrenal hyperplasia, Long-read sequencing, Comparative analysis, Clinical utility, Precise diagnosis

†Yunpeng Wang, Gaohui Zhu and Danhua Li contributed equally to this work.

*Correspondence:
Jing Zhu
jingzhu@cqmu.edu.cn
Min Zhu
zhumin323@hotmail.com
Full list of author information is available at the end of the article

Introduction

Congenital adrenal hyperplasia (CAH) is one of the most common autosomal recessive disorders caused by pathogenic variants in genes involved in adrenal steroidogenesis [1]. CAH ranges from the severe classic form to the asymptomatic non-classic form (NC). The classic CAH is subclassified into the potentially life-threatening salt-wasting (SW) form and the less severe simple-virilizing (SV) form. The frequency of the classic form in most populations varies between one in 10,000 and one in 20,000 [2–4], and NC form affects between one in 200 and one in 2000.

Variants in the *CYP21A2* gene result in steroid 21-hydroxylase deficiency (21-OHD), accounting for more than 95% of patients with CAH [2, 5]. *CYP21A2* is located on chromosome 6p21.33. The *CYP21A2* and *CYP21A1P* share 98% nucleotide sequence identity [6] and are aligned in tandem to constitute RCCX module, which contains the *RP* (also called *STK19* encoding a serine/threonine protein kinase), *C4* (encoding complement factor C4), *CYP21* (encoding steroid 21-hydroxylase) and *TNX* genes (encoding glycoprotein tenascin-X) [7–9]. A normal haplotype consists of two RCCX modules [10]. Single-nucleotide and insertion/deletion variants (SNVs/indels), as well as intergenic variants, resulted from microconversions between *CYP21A2* and *CYP21A1P* [11]. To date, more than 300 SNVs/indels have been reported [11, 12]. Unequal crossover between RCCX modules during meiosis generates large rearrangements leading to copy number variation (CNV) of RCCX including monomodular, bimodular and trimodular structure [9]. The 30 kb Deletion chimeras *CYP21A1P/A2* and *TNXA/B* in monomodular structure account for approximately 25%−30% of CAH [13]. Nine types of 30 kb deletion chimeras *CYP21A1P/A2* are classified into 7 classic SW forms (CH-1, CH-2, CH-3, CH-5, CH-6, CH-7 and CH-8) and 2 attenuated SV forms (CH-4 and CH-9) [14]. The 30 kb deletion chimeras *TNXA/B* are classified into three types (CH-1 to CH-3) [15]. The classification of chimeras contributes to the explanation of genotypes to phenotypes [16, 17]. Duplicated chimeras *CYP21A2/A1P* and *TNXB/A* in trimodular structure are also caused by unequal crossover, with the frequency of less than 1% [18, 19]. In addition to CAH patients caused by 21-OHD, there are other rare forms of CAH. Variants in the *CYP11B1* gene [20, 21], *HSD3B2*, *CYP17A1*, *CYP11A1*, *POR*, and *StAR* involved in steroid biosynthesis could also cause CAH [3].

Genetic testing of CAH plays a crucial role in precise diagnosis and disease management. The high homology between *CYP21A2* and *CYP21P1* makes genotyping challenging due to the complexity of various variants. Multiplex ligation-dependent probe amplification (MLPA) combined with Sanger sequencing is commonly used to identify CNV and SNVs/indels of *CYP21A2* [16, 22, 23]. However, MLPA depends on specific probes and cannot detect variants outside the target probe. Sanger sequencing is laborious and time-consuming.

Recently, next-generation Sequencing (NGS) such as gene panels, exome sequencing, genome sequencing have been adopted in genetic diagnosis [24–26]. However, the short reads (typically 100–300 bp) of NGS cannot be correctly aligned with the repetitive regions of the reference genome, making it difficult to detect the variants in highly homologous genes (*CYP21A2* and *CYP21A1P*). The hybrid approach combining PCR-based NGS, MLPA and Sanger sequencing was employed in the diagnosis of CAH [27–29]. Meanwhile, Long-read Sequencing (LRS) is a promising approach for many unsolved genetic disorders and has already been used for the precise diagnosis of CAH [30–32]. A comparative analysis between the targeted LRS (T-LRS) assay and the hybrid approach has not yet been conducted. Additionnaly, large-scale cohorts have not yet been enrolled in clinical testing.

In this study, the clinical efficacy and benefits of T-LRS were assessed in 322 probands clinically diagnosed with CAH. The retrospective study used the hybrid approach as a control method, while the prospective study focused on evaluating diagnostic efficacy of T-LRS, revealing the accurate genotypes of study cohort and assessing the clinical utility of T-LRS in CAH patients.

Materials and methods

Study participants and design

A total of 562 participants, including 322 probands with clinically diagnosed CAH and 240 family members, were enrolled from Affiliated Children's Hospital of Chongqing Medical University in China from January 2017 to July 2023. The recruited probands were diagnosed as CAH according to the clinical symptoms and genetic testing using control method in retrospective study. The inclusion criteria are as follows: (1) Persistent increase in blood 17-OHP concentration. The cut-off value was 10.5 nmol/L. (2) Elevated levels of blood androstenedione and dehydroepiandrosterone. (3) Elevated or normal levels of adrenocorticotropic hormone (ACTH). (4) Salt-wasting type shows decreased blood sodium and increased blood potassium levels. (5) Clinical manifestations include skin pigmentation and clitoral enlargement in female infants. (6) Genetic testing reveals specific mutations. A diagnosis can be made based on criteria (1), (2), and (3) or on (1) and (5).

The symptoms of SW form of CAH are as follows: Complete 21-OHD deficiency results in severe salt loss. Patients suffer from significant deficiencies in cortisol and aldosterone, along with excessive adrenal

androgen secretion. Symptoms of salt loss and adrenal cortex dysfunction may appear in the neonatal period, presenting as failure to gain weight gain or weight loss, persistent vomiting, diarrhea, weight loss, and difficult-to-correct hyponatremia, hyperkalemia, and metabolic acidosis. The symptoms of SV form of CAH are as follows: Incomplete 21-OHD deficiency allows for 1–2% of in vitro activity. This type primarily manifests with symptoms and signs of elevated adrenal androgen levels but without salt-wasting. SV patients may show peripheral precocious puberty, accelerated growth, and advanced bone age, while female patients often exhibit masculinization of external genitalia. The symptoms of NC form of CAH patients mainly present with hyperandrogenic symptoms, specifically manifested as early onset of pubic hair, hirsutism, acne and other Polycystic ovary syndrome (PCOS) like symptoms. In addition, the affected children often exhibit accelerated growth and advanced bone age.

In our cohort, we have one female patient with 11β-hydroxylase (*CYP11B1*) deficiency (11-OHD) who was first admitted at the age of one year due to abnormal external genitalia. Laboratory results showed a renin level below 0.5 μIU/ml and a 17-OHP level greater than 60.6 nmol/L and no electrolyte disorder. In contrast, patients with 21-OHD often exhibit elevated or normal levels of renin. Therefore, based on the patient's clinical presentation and sequencing results, a diagnosis of 11-OHD was made. Additionally, our cohort includes a male patient with 3β-hydroxysteroid dehydrogenase type 2 (*HSD3B2*) deficiency, who was first admitted shortly after birth due to undermasculinization and accompanied by skin pigmentation. Laboratory results showed that the child has low sodium, high potassium, elevated ACTH, androstenedione, and dehydroepiandrosterone. Based on clinical manifestations and sequencing results, a diagnosis of 3β-hydroxysteroid dehydrogenase type 2 deficiency was confirmed. Consent for genetic testing was obtained from each individual or legal guardians. This study was approved by the ethics board of Children's Hospital of Chongqing Medical University (Approval number: 2023165).

The flowchart of this study is illustrated in Fig. 1A. Initially, retrospective cohort studies were conducted to evaluate the results of genetic testing between T-LRS and control method. 96 probands (95 families) exhibiting clinical features of CAH were recruited for retrospective studies. Subsequently, prospective studies were also performed. 226 probands with clinical feature of CAH and 49 relatives were enrolled to assess the effect of T-LRS. Genomic DNA was extracted using the DNease Blood & Tissue Kit (Qiagen). The probands with suspected CAH were diagnosed by clinical presentation such as altered steroid hormone profiling, salt loss, abnormal sex development.

The combined assay used as control method

The hybrid approach employed in this study was a combined assay consisting of three methods including long-rang PCR-based NGS, MLPA and Sanger sequencing.

For the detection of SNVs/Indels in the functional gene *CYP21A2*, NGS based on long-range PCR was utilized. Specific primers were designed to amplify the entire *CYP21A2* gene, excluding interference from *CYP21A1P*. The sequences of primers used for long-range PCR are listed in Table S4. The resulting amplicon was applied to generate DNA libraries. Sequencing was performed on an Illumina NextSeq 550 following the manufacturer's protocol. Raw data for the targeted genes were aligned with the reference human genome sequence (GRCh38) using the Burrows-Wheeler alignment (version 0.6) tool [33].

To identify deletion chimeras, duplicated chimeras and specific SNVs/Indels, MLPA was employed using the SALSA MLPA probemix kit P050-C1 CAH (MRS Holland, the Netherlands) was used to detect CNV and gene conversions in the *CYP21A* and *TNX* genes [34]. Additionally, Sanger sequencing was used to confirm the identified SNVs/Indels.

Targeted long-read sequencing and data analysis

The 7.0–8.5 kb DNA fragments of *CYP21A2*, *CYP21A1P*, *CYP11B1*, *CYP17A1*, *HSD3B2*, and *StAR* genes were amplified using 6 locus-specific primer pairs though long-range PCR, as reported previously [32]. Briefly, purified PCR products underwent an end repair and ligation reaction to add barcoded adaptors and construct the single-molecule real-time (SMRT) bell library using the Sequel Binding Kit 2.0 and Internal Control Kit 1.0 (Pacific Biosciences). Primed DNA-polymerase complexes were loaded onto SMRT cells and sequenced with Sequel II Sequencing Kit 2.0 on the Sequel II platform (Pacific Biosciences).

The filtered CCS reads were obtained and then aligned to the human GRCh38/hg38 genome sequences using the SMRT Link analysis software suite (PacBio). FreeBayes 1.3.4 was employed for the detection of SNVs/Indels. The classification of variants were done according to ACMG guidelines [35–37]. The validation of discordant genotypes detected by T-LRS and control method was performed by Sanger sequencing.

Additional primer pairs covering *POR* and *CYP11A1* genes were also designed for the genetic diagnosis of samples with VUS.

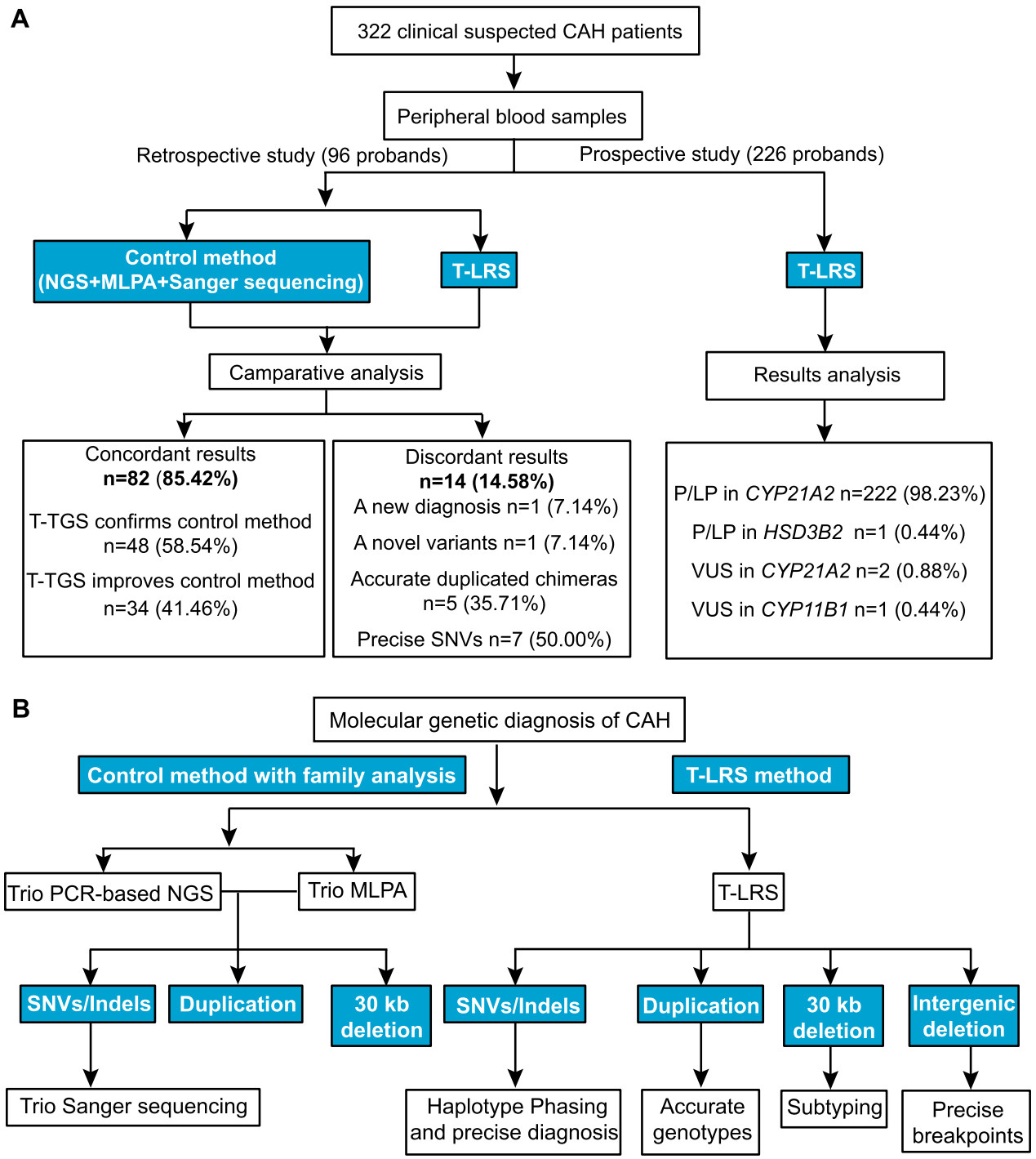

Fig. 1 Flow chart of comparative analysis and molecular genetic diagnosis and of control method and T-LRS strategies. **A** The flow chart of study design and comparative analysis of control method and T-LRS. Pathogenicity is evaluated according to ACMG guidelines. LP, Likely pathogenic; P, Pathogenic; VUS, Variant of uncertain significance. Probands carrying VUS presented evident clinical phenotypes of CAH. **B** The genetic testing chart of control method and T-LRS

Results

Patient characteristics

During the study period, a total of 562 participants including 322 affected individuals (54% males) and 240 family members met the inclusion criteria and were included in the final analysis. Among them, a total of

287 participants, including 96 probands and 191 family members from 95 families, were enrolled in a retrospective study (Table S1 and Table 1), while a total of 275 participants, including 226 probands and 49 relatives, were enrolled in a prospective study (Table S2 and Table 1). In this retrospective study, the genotypes of 287 participants were detected using both the control method and T-LRS (Table S1 and Table 1). The efficacy of T-LRS was evaluated through a comparative analysis between the two methods. A prospective study was subsequently conducted to accurately determine the genotypes of the 275 participants exclusively using T-LRS, in order to evaluate its clinical utility in CAH patients (Table S2 and Table 1).

In this study, the majority of probands (177) were diagnosed at birth due to clinical symptoms. This was followed by 48 probands diagnosed during preschool and 44 probands diagnosed during their school years. Only 9 out of 322 probands were diagnosed during adolescence (Table 1). All probands had onset before 18 years old, with median (minimum–maximum) 17-OHP value of 75.75 (11.49–907.8) nmol/L (Table 1).

A comparative analysis between T-LRS and control method in retrospective study

Two approaches, T-LRS without family analysis and control method with family analysis, were employed to conduct the molecular genetic testing and genotype the 96 probands and/or 191 family members. Based on the results of genetic testing, T-LRS exhibited concordance results with control method in 82 (85.42%) patients and discordant results in 14 (14.58%) patients (Fig. 1A). The genetic testing flowchart of control method and T-LRS was shown in Fig. 1B. The 82 patients with concordant results were further subclassified into two groups: (i) T-LRS without trio family analysis confirmed genotypes detected by control method with trio family analysis in 48 of 82 probands (58.54%) (Table S1). For instance, a compound heterozygote, one allele with c.1069C>T and the other allele with c.955C>T, were detected by T-LRS without family analysis in the case of proband P13. This demonstrated that T-LRS could directly obtain haplotype phasing and determine the cis/trans configuration (A); (ii) T-LRS improved outcomes determined by control method through identifying the breakpoints and subtypes of deletion chimeras in 34 of 82 probands (41.46%) (Table 2). For instance, a deletion chimera was identified and subclassified as *CYP21A2* CH-1, with the breakpoint marked by a blue box in Fig. 2B in proband P24. These results demonstrated that T-LRS can identify the breakpoints and subtypes of deletion chimeras in a single test, improving our understanding of CAH.

It is noteworthy that genotypes of 14 out of 96 probands (14.58%) were corrected by T-LRS (Table 3) and classified into five groups: (i) A new diagnosis was performed by T-LRS. The genetic diagnosis for proband P69 from family F68 was contradictory between T-LRS and control method. Initially, proband P69 was misdiagnosed by control method with a genotype of heterozygote for the c.293-13C>G variant. Actually, the result of T-LRS indicated that proband P69 was actually a compound heterozygote with c.293-13C>G on one allele and c.1306C>T (NC phenotype) on the other allele. These findings were confirmed though Sanger sequencing (Fig. 2C and Table 3). T-LRS improved the new clinical diagnosis and identified the missing disease-causing variants. (ii) A 4141 kb novel deletion resulting in a contiguous deletion of *CYP21A2* and *TNXB* was established by T-LRS in proband P17, initially detected

Table 1 Demographic characteristics of the 322 probands

Characteristics	Total (n)	Restropective study (n)	Prospective study (n)
Total number of probands	322	96	226
Total number of samples	562	287	275
Sex			
Male	175 (54%)	57 (59%)	118 (52%)
Female	147 (46%)	39 (41%)	108 (48%)
Age at onset			
Newborn	176	54	122
1–12 months (infant)	29	12	17
1–3 years (toddler)	16	6	10
3–6 years (preschool)	48	21	27
6–12 years (school age)	44	3	41
12–18 years (adolescence)	9	0	9
Median 17-OHP at diagnosis (minimum, maximum)	75.75 (11.49, 907.8)	75.75 (11.49, 907.8)	75.75 (11.9, 907.8)

Table 2 The genotypes of 33 probands were improved by T-LRS

No	Family	Proband	Genotype detected by T-LRS	Genotype detected by control method	The breakpoint of deletion chimeras
1	F2	P2	*CYP21A1P/A2*-CH-7/c.518T>A	Exon 1–6 deletion/c.518T>A	chr6:32041679–32042412
2	F3	P3	*CYP21A1P/A2*-CH-3/CYP21A1P/A2-CH-3	Exon 1–7 deletion/Exon1-7 deletion	chr6:32041679–32042412/ chr6:32041679–32042412
3	F4	P4	*TNXA/B*-CH-1/c.293-13C>G	Exon 1–7 deletion/c.293-13C>G	chr6:32044184–32045077
4	F6	P6	*CYP21A1P/A2*-CH-5/c.518T>A	Exon 1–7 deletion/c.518T>A	chr6:32040216–32040421
5	F10	P10	*TNXA/B*-CH-1/c.518T>A	Exon 1–7 deletion/c.518T>A	chr6:32044184–32045077
6	F18	P18	*TNXA/B*-CH-2/c.518T>A, c.923dup	Exon 1–7 deletion/c.518T>A, c.923dup	chr6:32042509–32043712
7	F21	P21	*CYP21A1P/A2*-CH-1/c.293-13C>G	Exon 1–3 deletion/c.293-13C>G	chr6:32039143–32039426
8	F22	P22	*TNXA/B*-CH-1/c.293-13C>G	Exon 1–3 deletion/c.293-13C>G	chr6:32044184–32045077
9	F24	P24	*CYP21A1P/A2*-CH-1/c.293-13C>G	Exon 1–3 deletion/c.293-13C>G	chr6:32039143–32039426
10	F29	P30	*CYP21A1P/A2*-CH-1/c.518T>A	Exon 1–3 deletion/c.518T>A	chr6:32039143–32039426
11	F34	P35	*CYP21A1P/A2*-CH-1/c.293-13C>G	Exon 1–3 deletion/c.293-13C>G	chr6:32039143–32039426
12	F35	P36	*CYP21A1P/A2*-CH-1/CYP21A1P/A2-CH-1	Exon 1–3 deletion/Exon1-3 deletion	chr6:32039143–32039426/ chr6:32039143–32039426
13	F38	P39	*CYP21A1P/A2*-CH-8/c.293-13C>G	Exon 1–7 deletion/c.293-13C>G	chr6:32041679–32042412
14	F39	P40	*CYP21A1P/A2*-CH-1/c.293-13C>G	Exon 1–3 deletion/c.293-13C>G	chr6:32039143–32039426
15	F43	P44	*TNXA/B*-CH-3/c.293-13C>G	Exon 1–7 deletion/c.293-13C>G	chr6:32041679–32042412
16	F48	P49	*CYP21A1P/A2*-CH-1/c.293-13C>G	Exon 1–3 deletion/c.293-13C>G	chr6:32039143–32039426
17	F51	P52	*TNXA/B*-CH-1/c.293-13C>G	Exon 1–7 deletion/c.293-13C>G	chr6:32044184–32045077
18	F56	P57	*TNXA/B*-CH-1/c.293-13C>G	Exon 1–7 deletion/c.293-13C>G	chr6:32044184–32045077
19	F64	P65	*CYP21A1P/A2*-CH-2/c.293-13C>G	Exon 1–4 deletion/c.293-13C>G	chr6:32039548–32039802
20	F66	P67	*CYP21A1P/A2*-CH-8/c.518T>A	Exon 1–7 deletion/c.518T>A	chr6:32041679–32042412
21	F69	P70	*CYP21A1P/A2*-CH-1/c.332_339del	Exon 1–3 deletion/c.332_339del	chr6:32039143–32039426
22	F70	P71	*CYP21A1P/A2*-CH-2/c.332_339del	Exon 1–4 deletion/c.332_339del	chr6:32039548–32039802
23	F73	P74	*CYP21A1P/A2*-CH-1/c.518T>A	Exon 1–3 deletion/c.518T>A	chr6:32039143–32039426
24	F74	P75	*TNXA/B*-CH-3/TNXA/B-CH-1	Exon 1–7 deletion/Exon 1–6 deletion	chr6:32042415–32042485/ chr6:32044184–32045077
25	F77	P78	*TNXA/B*-CH-1/c.293-13C>G	Exon 1–7 deletion/c.293-13C>G	chr6:32046069–32046114
26	F78	P79	*TNXA/B*-CH-1/c.518T>A	Exon 1–7 deletion/c.518T>A	chr6:32044184–32045077
27	F82	P83	*CYP21A1P/A2*-CH-1/c.292+1G>A	Exon 1–3 deletion/c.292+1G>A	chr6:32039143–32039426
28	F83	P84	*CYP21A1P/A2*-CH-8/c.293-13C>G	Exon 1–7 deletion/c.293-13C>G	chr6:32041524–32041574
29	F84	P85	*TNXA/B*-CH-1/c.1148_1149del	Exon 1–7 deletion/c.1148_1149del	chr6:32044184–32045077
30	F85	P86	*TNXA/B*-CH-1/c.293-13C>G	Exon 1–7 deletion/c.293-13C>G	chr6:32044184–32045077
31	F87	P88	*CYP21A1P/A2*-CH-1/c.293-13C>G, c.518T>A	Exon 1–3 deletion/c.293-13C>G, c.518T>A	chr6:32039143–32039426
32	F88	P89	*CYP21A1P/A2*-CH-1/CYP21A1P/A2-CH-1	Exon 1–3 deletion/Exon 1–3 deletion	chr6:32039143–32039426/ chr6:32039143–32039426
33	F89	P90	*CYP21A1P/A2*-CH-1/c.1070G>A	Exon 1–3 deletion/c.1070G>A	chr6:32039143–32039426
34	F92	P93	*TNXA/B*-CH-1/c.1451G>C	Exon 1–7 deletion/c.1451G>C	chr6:32044177–32044184

as exon 3 deletion by control method. This intergenic deletion was confirmed by Sanger sequencing (Fig. 2D). The sequences of primers used in the confirmation process are listed in Table S4. The results of family analysis indicated that the novel deletion was inherited from his mother and confirmed by Sanger sequencing (Fig. 2D). Mis-alignment and micro-conservation during meiosis of homologous genes (*CYP21A1P* and *CYP21A2*) were responsible for this large deletion variants; (iii) Five out of the 14 probands including P54, P50, P55, P61 and P46 with duplicated chimeras, initially not precisely identified by control method, were detected by T-LRS (Table 3). For instance, in P54, one allele carried deletion chimera *TNXA/B* CH-1, and the other allele had two SNVs (c.923dup and c.955C>T) on both *CYP21A2* gene and duplicated chimera *CYP21A2/A1P*. The breakpoint was established and marked by a red box in Fig. 3A. Haplotype of *TNXA/B* chimera with a

Fig. 2 IGV plots of representative samples based on the comparative analysis. **A** The genetic results of control method were comfirmed by T-LRS. **B** T-LRS improved the genetic diagnosis by determining the subtypes and breakpoints of chimeras. **C** A new diagnosis was improved by T-LRS. **D** A novel 4141 kb deletion was identified by T-LRS and the family analysis was conducted. The variants were confirmed by Sanger sequencing

CYP21A2 and *TNXB* deletion confused duplicated chimera with deletion chimera in MLPA of P54 (Fig. S1). If an individual carries a duplicated *CYP21A2* on one allele and a deletion on the other allele, as seen in this case, the duplicated *CYP21A2* could conceal the presence of the deletion. This issue must rely on lineage analysis; otherwise, it cannot be resolved. However, T-LRS can directly identify the precise genotype in this circumstance. (iv) T-LRS conducted precise diagnosis of SNVs compared with the control method (Table 3). In 2 probands, SNV (c.−126C>T) leading to non-classical CAH could not be detected by control method. For example, in proband P45, the SNV (c.−126C>T) was identified by T-LRS (Fig. 3B). In proband P66 from

Table 3 The genotypes of 14 probands were corrected by T-LRS

No	Family	Proband	T-LRS	Control method	Benefits of T-LRS
1	F68	P69	c.293-13C>G/c.1306C>T	c.293-13C>G/+	A new diagnosis improved by T-LRS
2	F17	P17	chr6:32038800-32042941_del/c.293-13C>G	Exon 3 deletion, c.293-13C>G/c.293-13C>G	A novel variant with a 4141 kb deletion detected by T-LRS
3	F53	P54	*TNXA/B*-CH-1/c.923dup, c.955C>T, *CYP21A2/A1P with* c.923dup, c.955C>T	Exon 1–7 deletion/c.923dup,c.955C>T	Detecting accurate genotype of duplicated chimera
4	F49	P50	*TNXA/B*-CH-2/c.518T>A, *CYP21A2/A1P*	Exon 1–7 deletion/c.518T>A	Detecting accurate genotype of duplicated chimera
5	F54	P55	*TNXA/B*-CH-2/c.1024C>T, *CYP21A2/A1P*	Exon 1–7 deletion/c.1024C>T	Detecting accurate genotype of duplicated chimera
6	F60	P61	*CYP21A1P/A2*-CH-1/c.844G>T, c.923dup, c.955C>T, *CYP21A2/A1P* with c.844G>T, c.923dup, c.955C>T	Exon 1–3 deletion/*CYP21A2/A1P*, c.844G>T, c.923dup, c.955C>T	Detecting accurate genotype of duplicated chimera
7	F45	P46	c.518T>A/c.844G>T, c.923dup, c.955C>T, *CYP21A2/A1P* with c.844G>T, c.923dup, c.955C>T	c.518T>A/*CYP21A2/A1P*, c.844G>T, c.923C>T, c.955C>T	Detecting accurate genotype of duplicated chimera
8	F44	P45	c.−126C>T, c.332_339del/c.955C>T	c.332_339del/c.955C>T	Precise diagnosis of SNV c.−126C>T
9	F94	P95	*TNXA/B*-CH-1/c.−126C>T, c.−113G>A	Exon 1–7 deletion/c.−113G>A	Precise diagnosis of SNV c.−126C>T
10	F65	P66	c.518T>A/c.1451_1452delinsC	c.518T>A/c.1455del	Precise diagnosis of SNV c.1451_1452delinsC
11	F15	P15	E6_cluster(c.710T>A,c.713T>A,c.719T>A), c.844G>T, c.923dup, c.955C>T, c.1069C>T/c.518T>A	Exon 4–7 deletion/c.518T>A	Precise diagnosis of Continuous SNVs
12	F23	P23	E6_cluster(c.710T>A, c.713T>A, c.719T>A), c.844G>T, c.923dup, c.1069C>T/c.202+1G>A	Exon 6–7 deletion/c.202+1G>A	Precise diagnosis of Continuous SNVs
13	F36	P37	c.518T>A, E6_cluster(c.710T>A, c.713T>A, c.719T>A), c.844G>T, c.923dup, c.1069C>T/c.293-13C>G	Exon 4–7 deletion, c.1069C>T/c.293-13C>G	Precise diagnosis of Continuous SNVs
14	F90	P91	E6_cluster(c.710T>A, c.713T>A, c.719T>A), c.844G>T, c.923dup, c.955C>T, c.1069C>T, c.1451_1452delinsC/c.518T>A	Exon 6–7 deletion/c.518T>A	Precise diagnosis of Continuous SNVs

family F65, the SNV c.1451_1452delinsC, rather than c.1455del, was identified using T-LRS and confirmed though Sanger sequencing (Fig. 3C). Four probands with contiguous SNVs were incorrectly reported as deletion chimeras by the control method. This type of discordant result is more likely attributed to the recognition process of the genetic testing results by the control method. In the control method, long-rang PCR-based NGS could identify the continuous SNVs but cannot detect copy number variants, so MLPA was used to detect the copy number variants. When long-rang PCR-based NGS identifies continuous SNVs, while MLPA detects deletion, as in these cases, the variants are recognized as a deletion.

In the retrospective study, the diagnostic yield for patients detected using the control method was 97.92% (94 out of 96 patients). In comparison, T-LRS led to one additional diagnosis, resulting in a diagnostic yield 98.96% (95 of 96 patients). The false negative rate for the control method was 1.04% (1 of 96 patients).

Detection of genotypes in 226 probands in prospective study

Based on the results of retrospective study mentioned above, the detection of genotypes by T-LRS was proven to be effective and convenient. The prospective study exclusively employed T-LRS to detect the variants in 226 probands and 49 relatives, reassessing the application of T-LRS for the genetic diagnosis of CAH (Table S2). The genetic testing results indicated that 223 (223/226, 98.67%) probands with P/LP variants and 3 (3/226, 1.33%) probands with VUS, representing clinical features of CAH (Fig. 1A). The genetic testing flowchart of T-LRS was shown in Fig. 1B. Among the probands, 98.23% (222/226) had biallelic pathogenic variants of *CYP21A2* gene, leading to 21-OHD. An extremely rare case of classic form, caused by biallelic variants in the *HSD3B2*, was identified in one patient (1/226, 0.44%). Three probands including P97, P114 and P197 carried VUS, with P97 and P114 having compound heterozygote involving VUS on one alleles and

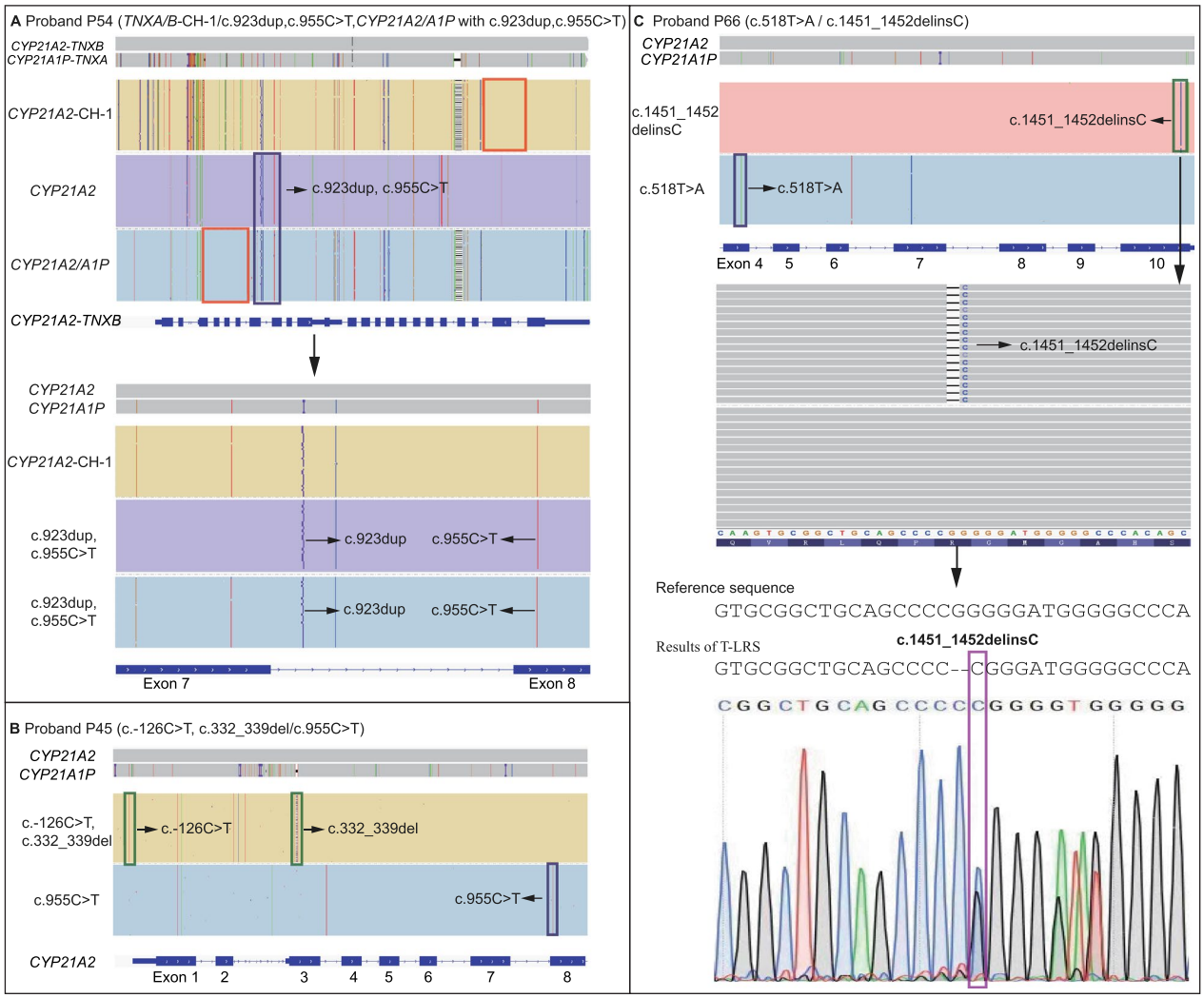

Fig. 3 The genotypes of probands with duplicated chimeras and SNVs/Indels corrected by T-LRS. **A** Duplicated chimera with SNVs/Indels was identified by T-LRS. **B** Variants c.−126C>T in the promoter region of *CYP21A2*, not detected by control method, were identified by T-LRS. **C** Proband P66 had the c.1451_1452delinsC, not the c.1455del identified by T-LRS. The variant was confirmed by Sanger sequencing

different variants on the other allele in the *CYP21A2* gene, and Proband P197 harboring a homozygote for VUS in the *CYP11B1* gene, whose genotype is unpublished previously (Table S2).

Among the 222 probands with biallelic pathogenic variants of the *CYP21A2* gene, 126 affected persons were compound heterozygotes with same (37 probands) or different (90 probands) SNVs/Indels on each allele. 85 patients had deletion chimeras on one allele and SNVs/Indels on the other allele, while five probands harbored deletion chimeras on both alleles. Additionally, five probands carried duplicated chimeras. Above all, molecular genetic diagnosis using T-LRS precisely detected the disease-causing variants and VUS in CAH patients, demonstrating a good genotype–phenotype correlation.

The spectrum of genetic variants identified by T-LRS in all the probands

In total, 322 affected individuals underwent genetic diagnosis using T-LRS in this study. Among the 322 probands, 99.38% (320/322) had biallelic variants due to 21-OHD. Only 0.31% probands (1/322) harbored different SNVs on each allele in the *HSD3B2* gene, and one proband (1/322) carried biallelic VUS in the *CYP11B1* gene (Table S3). 57.19% out of 21-OHD patients (183/320) had SNVs/Indels on both alleles in *CYP21A2* gene, and their genotypes were classified into 72 different types. 35.94% out of 21-OHD patients (115/320) had deletion monoallelic chimeras with 51 different types, while 3.44% of 21-OHD patients (11/320) had biallelic deletion chimeras of 9 types. 0.31% of 21-OHD

patients (1/320) had a novel monoallelic 4141 kb deletion, and 3.13% of 21-OHD patients (10/320) had duplicated chimeras of 10 types (Table S3).

Patients with monoallelic and biallelic *TNXA/B* are more likely to have Ehlers–Danlos syndrome (EDS), which can affect musculoskeletal, dermatological, cardiovascular and gastrointestinal systems (Table S5). A total of 59 patients with biallelic or monoallelic *TNXA/B* were evaluated. The most common presentation was delayed gastric emptying, with the frequency of 28.81% (17/59), followed by recurrent joint dislocations (11.86%, 7/59).

Alleles detected by T-LRS were further analyzed (Table 4). In total, 486 alleles with SNV/indels of 65 types (75.47%, 486/644), 143 alleles with deletion chimeras of 10 types (22.20%, 143/644), 10 alleles with deplicated chimeras (1.55%, 10/644), and one allele with a novel 4141 kb deletion in the *CYP21A2* gene were identified. Two alleles with SNVs in the *HSD3B2* gene were detected, along with two alleles with VUS in the *CYP11B1* gene were identified (Table 4). The most frequent SNV was c.293-13C>G (32.76%, 211/644), followed by c.518T>A (15.53%, 100/644) and c.1069C>T (4.66%, 30/644). The most frequent deletion chimera was *CYP21A2*-CH-1 (7.30%, 47/644), followed by *TNXB* CH-1 (5.43%, 35/644) and *TNXB* CH-2 (3.11%, 20/644). The most frequent duplicated chimera was a *CYP21A2* gene with c.844G>T, c.923dup and c.955C>T plus duplicated *CYP21A2/A1P* (0.78%, 5/644) (Table 4). The observed frequency of duplication haplotype was 1.55%.

The spectrum of marker pathogenic variants derived from *CYP21A1P* in the deletion chimeras was displayed in Fig. 4. Not only conserved but also non-conserved variants were found in samples with deletion chimeras. The most non-conserved variant was c.1069C>T with 66.2% variant frequency, followed by c.92C>T (78.87%), c.12514G>A (76.92%), c.844G>T (75.9%) and c.12524G>A (73.85%).

Detection of genotypes with VUS

In total, 4 probands carried VUS (Table 5). Among these, three probands were compound heterozygotes with a VUS on one allele and a pathogenic/likely pathogenic variant on other allele in the *CYP21A2* gene. One proband carried a novel genotype, which had not been published previously, with a homozygote VUS c.1145T>C in *CYP11B1*. This patient presented increased concentration of 17-OHP (60.6 nmol/L) and clitoromegaly (Table 5). Affected female individuals P97, P114, and P197 all had atypical genitalia in neonates or infants with a 46, XX karyotype. The remaining 1 male proband exhibited signs of androgen excess including acne, premature pubarche, and accelerated growth during childhood. Increased 17-OHP concentrations (ranging from 34.64 to 451.2 nmol/L), the primary symptoms of CAH, were also observed.

Discussion

This study demonstrated the efficacy and benefits of T-LRS, primarily in comparison with the combined strategy that includes long-range NGS and MLPA, and evaluated the clinical utility of T-LRS in a large cohort of 322 cases. In the retrospective study, T-LRS can identify pathogenic variants that have been missed by control method, providing timely treatment to patients. Moreover, a novel deletion was detected by T-LRS. Prospective study was also designed to further assess clinical application of T-LRS in CAH. Not only a series of variants in *CYP21A2* were detected, but also other extremely rare variants in *HSD3B2* and *CYP11B1* were identified. This study is the first time that the efficacy and benefits of T-LRS were accessed by comparative analysis between T-LRS and control method in a large cohort of 322 probands.

The 85.42% concordant results were found between T-LRS and control method in the retrospective study. The T-LRS was demonstrated to detect different types of *CYP21A2* variants accurately and identify the breakpoints of deletions precisely [14, 38]. Patients with monoallelic *TNXA/B* are more likely have EDS due to Tenascin-X deficiency [15]. The identification of chimeras *TNXA/B* reminds the clinicians to assess clinical features of EDS for precise patient therapy. It is of note that the junction sites of chimeras could also be determined by Sanger sequencing using specific primers, which were labor-intensive and difficult to scale up [22].

Discordant results (14.58%) were detected between T-LRS and control method. Originally, proband P69 was misdiagnosed by control method. T-LRS successfully diagnosed this CAH patient through detection of causative variant with NC CAH. One possible reason is that short-read sequencing-based NGS, although accurate, has GC bias in sequencing [30, 39, 40]. Additionally, T-LRS conducted precise diagnosis of SNVs compared with the control method. The variant c.−126C>T was likely missed by NGS due to an inadequate understanding of the regulatory landscape of genes, leading to its exclusion during interpretation of variants [41, 42]. The difficulty in detecting the variant c.1451_1452delinsC by NGS was likely due to the presence of repetitive GG and CC sequences, which pose a challenge for NGS. The variant c.1451_1452delinsC can also be represented as c.1451G>C and c.1455del. Thus, NGS identified only c.1455del, missing c.1451G>C. This is likely due to the challenges that short-read sequencing-based NGS struggles to reliably map sequence reads to repetitive regions, such as low-complexity regions enriched for GC- or AT-rich DNA, as in this case [30, 31, 40].

Table 4 Allele frequency of variants in the 322 probands

Gene	Type of variation	Nucleotide variation	Allele number	Frequency of alleles
CYP21A2	SNVs/Indels	66 types of alleles with SNVs/Indels in *CYP21A2*	486 (total)	75.47% (total)
	1	c.293-13C>G	211	211/644 (32.76%)
	2	c.518T>A	100	100/644 (15.53%)
	3	c.1069C>T	30	30/644 (4.66%)
	4	c.955C>T	20	20/644 (3.11%)
	5	c.332_339del	15	15/644 (2.33%)
	6	c.1451_1452delinsC	14	14/644 (2.17%)
	7	c.92C>T	6	6/644 (0.93%)
	8	c.292+1G>A	6	6/644 (0.93%)
	9	c.955C>T, c.1069C>T	6	6/644 (0.93%)
	10	c.923dup	5	5/644 (0.78%)
	11	c.844G>T, c.923dup, c.955C>T, c.1069C>T	4	4/644 (0.62%)
	12	c.293-13C>G, c.332_339del	4	4/644 (0.62%)
	13	c.1451G>C	3	3/644 (0.47%)
	14	c.1306C>T	2	2/644 (0.31%)
	15	c.169G>A	2	2/644 (0.31%)
	16	E6_cluster(c.710T>A,c.713T>A,c.719T>A)	2	2/644 (0.31%)
	17	c.1279C>T	2	2/644 (0.31%)
	18	c.1280G>A	2	2/644 (0.31%)
	19	c.1450dup	2	2/644 (0.31%)
	20	c.92C>T,c.293-13C>G,c.332_339del,c.518T>A, E6_cluster(c.710T>A,c.713T>A, c.719T>A),c.844G>T,c.923dup,c.955C>T,c.1069C>T	2	2/644 (0.31%)
	21	c.949C>T	2	2/644 (0.31%)
	22	c.518T>A,E6_cluster(c.710T>A,c.713T>A,c.719T>A),c.844G>T,c.923dup,c.1069C>T	2	2/644 (0.31%)
	23	c.1064G>A	1	1/644 (0.16%)
	24	E6_cluster(c.710T>A,c.713T>A,c.719T>A),c.844G>T,c.923dup,c.955C>T,c.1069C>T	1	1/644 (0.16%)
	25	E6_cluster(c.710T>A,c.713T>A,c.719T>A),c.844G>T,c.923dup,c.1069C>T	1	1/644 (0.16%)
	26	c.−126C>T,c.332_339del	1	1/644 (0.16%)
	27	c.518T>A,E6_cluster(c.710T>A,c.713T>A,c.719T>A),c.844G>T,c.923dup,c.955C>T	1	1/644 (0.16%)
	28	c.535G>A	1	1/644 (0.16%)
	29	E6_cluster(c.710T>A,c.713T>A, c.719T>A),c.844G>T,c.923dup,c.955C>T,c.1451_1452delinsC	1	1/644 (0.16%)
	30	E6_cluster(c.710T>A,c.713T>A,c.719T>A),c.797C>T,c.844G>T,c.923dup,c.955C>T,c.1069C>T,c.1451_1452delinsC	1	1/644 (0.16%)
	31	c.1273G>A	1	1/644 (0.16%)
	32	c.92C>T,c.518T>A,c.844G>T	1	1/644 (0.16%)
	33	c.−126C>T,c.−113G>A,c.518T>A	1	1/644 (0.16%)
	34	E6_cluster(c.710T>A,c.713T>A,c.719T>A),c.844G>T,c.923dup,c.955C>T,c.1069C>T,c.1451_1452delinsC	1	1/644 (0.16%)
	35	c.377C>G	1	1/644 (0.16%)

Table 4 (continued)

Gene	Type of variation	Nucleotide variation	Allele number	Frequency of alleles
	36	c.−126C>T,c.−113G>A,c.92C>T	1	1/644 (0.16%)
	37	c.−126C>T,c.−113G>A,c.−110T>C	1	1/644 (0.16%)
	38	c.433C>T	1	1/644 (0.16%)
	39	c.518T>A,E6_cluster(c.710T>A,c.713T>A,c.719T>A),c.844G>T,c.923dup	1	1/644 (0.16%)
	40	c.1218G>A	1	1/644 (0.16%)
	41	E6_cluster(c.710T>A,c.713T>A,c.719T>A),c.923dup,c.955C>T	1	1/644 (0.16%)
	42	c.1136T>A	1	1/644 (0.16%)
	43	c.1273_1277del	1	1/644 (0.16%)
	44	c.916G>T(VUS)	1	1/644 (0.16%)
	45	c.518T>A,c.923dup	1	1/644 (0.16%)
	46	c.844G>T	1	1/644 (0.16%)
	47	c.202+1G>A	1	1/644 (0.16%)
	48	c.1054G>A	1	1/644 (0.16%)
	49	c.923dup,c.955C>T	1	1/644 (0.16%)
	50	c.1148_1149del	1	1/644 (0.16%)
	51	c.293-13C>G,c.518T>A	1	1/644 (0.16%)
	52	c.1070G>A	1	1/644 (0.16%)
	53	c.−126C>T,c.−113G>A	1	1/644 (0.16%)
	54	c.431C>A (VUS)	1	1/644 (0.16%)
	55	c.275_276delGAinsTT (VUS)	1	1/644 (0.16%)
	56	c.923dup,c.955C>T,c.1069C>T	1	1/644 (0.16%)
	57	c.1063C>T	1	1/644 (0.16%)
	58	c.772_783del	1	1/644 (0.16%)
	59	c.518T>A,E6_cluster(c.710T>A,c.713T>A,c.719T>A), c.923dup	1	1/644 (0.16%)
	60	c.−126C>T,c.−113G>A,c.−110T>C,c.293-13C>G	1	1/644 (0.16%)
	61	c.308delinsAAA	1	1/644 (0.16%)
	62	c.874G>A	1	1/644 (0.16%)
	63	c.−126C>T,c.−113G>A,c.−110T>C,c.−103A>G,c.92C>T,c.293-13C>G,c.332_339del	1	1/644 (0.16%)
	64	c.740del	1	1/644 (0.16%)
	65	c.293-13C>G,c.1069C>T	1	1/644 (0.16%)
	66	c.293-13C>G,c.1451_1452delinsC	1	1/644 (0.16%)
	Deletion chimeras	10 types of alleles with deletion chimeras in *CYP21A2*	143 (total)	22.20% (total)
	1	*CYP21A1P/A2*-CH-1	47	47/644 (7.30%)
	2	*TNXA/B*-CH-1	35	35/644 (5.43%)
	3	*TNXA/B*-CH-2	20	20/644 (3.11%)
	4	*CYP21A1P/A2*-CH-8	10	10/644 (1.55%)
	5	*TNXA/B*-CH-3	9	9/644 (1.40%)
	6	*CYP21A1P/A2*-CH-3	6	6/644 (0.93%)

Table 4 (continued)

Gene	Type of variation	Nucleotide variation	Allele number	Frequency of alleles
	7	*CYP21A1P/A2*-CH-2	5	5/644 (0.78%)
	8	*CYP21A1P/A2*-CH-5	4	4/644 (0.62%)
	9	*CYP21A1P/A2*-CH-9	4	4/644 (0.62%)
	10	*CYP21A1P/A2*-CH-7	3	3/644 (0.47%)
	Duplicated chimeras	5 types of alleles with duplicated chimeras in *CYP21A2*	10 (total)	1.55% (total)
	1	c.844G>T, c.923dup, c.955C>T, *CYP21A2/A1P*	5	5/644 (0.78%)
	2	c.518T>A, *CYP21A2/A1P*	2	2/644 (0.31%)
	3	c.923dup, c.955C>T, *CYP21A2/A1P*	1	1/644 (0.16%)
	4	c.955C>T, *CYP21A2/A1P*	1	1/644 (0.16%)
	5	c.1024C>T, *CYP21A2/A1P*	1	1/644 (0.16%)
	A novel deletion	An allele with a novel deletion in *CYP21A2*	1 (total)	0.16% (total)
	1	chr6:32038800-32042941_del	1	1/644 (0.16%)
HSD3B2	SNVs/Indels	2 types of alleles with SNVs/Indels in *HSD3B2*	2 (total)	0.31% (total)
	1	c.776C>T	1	1/644 (0.16%)
	2	c.1003C>T	1	1/644 (0.16%)
CYP11B1	SNVs/Indels	One type of allele with SNVs/Indels in *CYP11B1*	2 (total)	0.31% (total)
	1	c.1145T>C (VUS)	2	2/644 (0.31%)

A novel intergenic 4141 kb deletion resulting in the partial deletion of *CYP21A2* and *TNXB* was identified by T-LRS, highlighting the ability to detect potential novel variant in T-LRS. Duplicated chimeras can be precisely diagnosed using T-LRS. Additionally, breakpoints of the duplicated chimeras were detected. Overall, T-LRS were proven to have higher resolution than control method though detecting missing variant, potential novel deletion and CNV of RCCX. An accurate diagnosis can result in better management of CAH and advanced therapeutics options.

Prospective study was designed to further assess clinical application of T-LRS in CAH. Variants in *CYP21A2, HSD3B2 and CYP11B1* were all identified by T-LRS. Alongside accurate diagnosis of duplicated chimeras, it is possible to calculate the actual frequency of trimodular pathogenic haplotypes harboring the c.955C>T (p.Gln319Ter) on a duplicated *CYP21A2* gene, with a frequency of 1.09% (7/644) in pathogenic alleles, accounting for 70% (7/10) in trimodular pathogenic haplotypes [19, 43]. A total of 1.55% of trimodular haplotypes were identified in this study. This more precise CAH diagnosis contributes towards the avoidance of false genotyping, evaluation of CNV distribution and variants' linkage disequilibrium.

In total, 4 probands of 322 probands carried VUS. A novel genotype with homozygote for a VUS c.1145T>C in *CYP11B1* was identified, and the clinical features were also described. This is the first report of this genotype and different from the case with a monoallelic VUS c.1145T>C published previously [44]. In addition to large databases with precise genotype to phenotype were accumulated by T-LRS testing, it is possible to evaluate the potential pathogenicity of VUS.

When this study was initiated, NGS technology was widely available in clinical use and less expensive than LRS. However, due to the high homology between *CYP21A2* and *CYP21A1P*, various disease-causing variants including SNVs/Indels, large rearranges and CNVs pose challenges in the control method. Moreover, trio family analysis was necessary to identify the haplotype phasing and cis/trans configuration of multiple SNVs/Indels, owing to the limitation of short reads. Trio studies required by control method cost more than T-LRS. Currently, the overall cost of T-LRS is less than $20 USD per test [32, 45, 46]. Whereas the cost of PCR-based NGS and MLPA was approximately $25 USD and $22.5 USD per test, respectively [45, 47, 48]. The cost per T-LRS test is lower or comparable to that of NGS and MLPA. However, NGS and MLPA need to be combined for diagnosis and often requiring trio studies, which increases the overall diagnostic cost for each family.

As mentioned above, the control method has limitations, such as inaccurate SNV detection, inaccurate identification of duplicated chimeras using MLPA [32, 49] and higher costs compared to T-LRS due to the

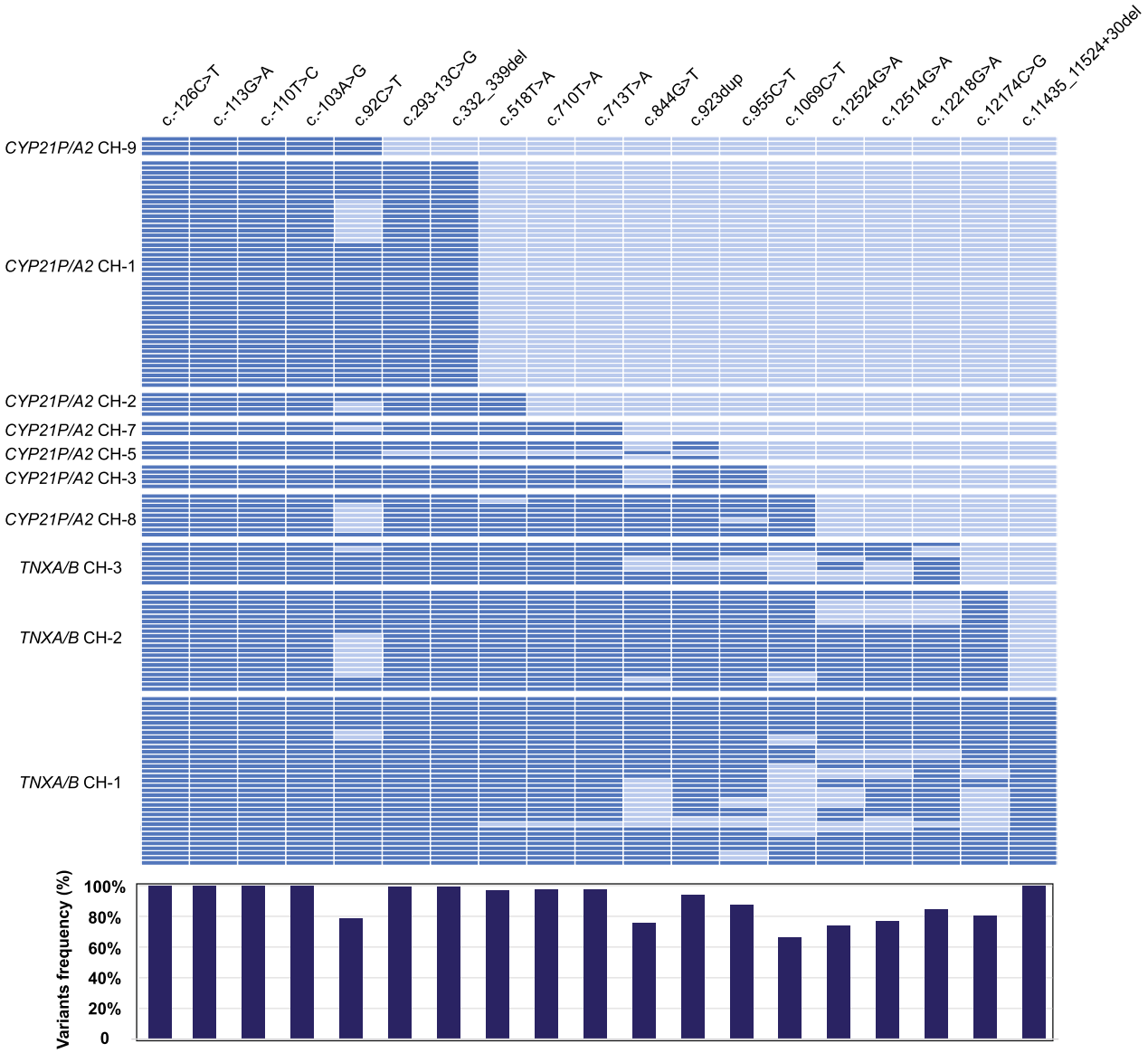

Fig. 4 The spectrum of pathogenic variants derived from *CYP21A1P* in the 7 types of *CYP21P/A2* deletion chimeras and 3 types of *TNXA/B* deletion chimeras in 322 probands. In the upper heatmap, row colors indicate the presence (dark blue) or absence (light blue) of pathogenic variants. Histogram at the bottom displays the variants frequency

need for family trio tests. LRS has already been used for the precise diagnosis of CAH [30–32]. This study evaluates the clinical efficacy and benefits of T-LRS in a large cohort of 322 probands with clinical presentations. It is suggested that T-LRS is a promising genetic test in CAH diagnosis. Considering the complex benefits and reduced cost of T-LRS, the substantial barrier to the widespread clinical use of T-LRS has been reduced. Accurate diagnosis of T-LRS can provide opportunities for timely intervention in CAH patients, advance the life-long management of this complex disease and promote our understanding of CAH. Available evidence strongly supports the use of the -LRS as the first-tier diagnostic test for CAH probands.

Table 5 The genotypes and clinical features in the individuals carrying the variants of uncertain significance

Genes	Proband	Sex	Age (years)	17-OHP	Symptoms	Genotype	Predicted protein change	Pathogenicity
CYP11B1	P197	Female	0.58	> 60.6	Clitoromegaly	c.1145T>C homo (VUS)	p.L382P	VUS
CYP21A2	P7	Male	6.5	34.64	B1PH2, acne, premature pubarche, accelerated childhood growth	c.1064G>A/c.916G>T (VUS)	p.V306F	VUS
CYP21A2	P97	Female	Newborn	451.2	B1PH1, PraderIII, hyperpigmentation, genital atypia, clitoromegaly	*CYP21A1P/A2*-CH-1/c.431C>A (VUS)	p.T144N	VUS
CYP21A2	P114	Female	Newborn	151.5	B1PH1, clitoromegaly	c.293-13C>G/c.275_276delGAinsTT (VUS)	p.R92I	VUS

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40246-024-00696-4.

Additional file 1.

Additional file 2.

Acknowledgements
Not applicable.

Author contributions
YW, GZ and DL have contributed equally to the article. They contributed to analyze data, draft the manuscript and revise it. YP, RL and TZ is responsible for the clinical studies. AM and LC are responsible for molecular genetic testing. JZ and MZ ensured the integrity of the entire study and reviewed the manuscript. All authors have read and approved the final manuscript.

Funding
The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the National Natural Science Foundation of China (Grant Number 81170723).

Availability of data and materials
All data generated during this study are included in this article.

Declarations

Ethics approval and consent to participate
This study was approved by the Ethics board of Children's Hospital of Chongqing Medical University (Approval number: 2023165). Informed written consent was obtained from all the subjects or their legal guardians.

Consent for publication
All authors have read and approved the final manuscript.

Competing interests
D.L, A.M. and J.L. are employees of Berry Genomics Corporation. The authors declare that they have no conflict of interest.

Author details
[1]Department of Endocrine and Metabolic Diseases, Children's Hospital of Chongqing Medical University, National Clinical Research Center for Child Health and Disorders, Ministry of Education Key Laboratory of Child Development and Disorders, Chongqing, China. [2]Chongqing Key Laboratory of Pediatrics, Children's Hospital of Chongqing Medical University, Chongqing, China. [3]Department of Pediatric Research Institute, Children's Hospital of Chongqing Medical University, National Clinical Research Center for Child Health and Disorders, Ministry of Education Key Laboratory of Child Development and Disorders, Chongqing Key Laboratory of Pediatrics, Chongqing, China. [4]Berry Genomics Corporation, Beijing 102200, China. [5]Chongqing Key Laboratory of Structural Birth Defect and Reconstruction, Chongqing, China.

Received: 10 September 2024 Accepted: 11 November 2024
Published online: 14 January 2025

References

1. Auer MK, Nordenström A, Lajic S, Reisch N. Congenital adrenal hyperplasia. Lancet (London, England). 2023;401(10372):227–44.
2. Speiser PW, Arlt W, Auchus RJ, Baskin LS, Conway GS, Merke DP, et al. Congenital adrenal hyperplasia due to steroid 21-hydroxylase deficiency: an endocrine society clinical practice guideline. J Clin Endocrinol Metab. 2018;103(11):4043–88.
3. Miller WL. Mechanisms in endocrinology: rare defects in adrenal steroidogenesis. Eur J Endocrinol. 2018;179(3):R125-r141.
4. Wilson RC, Nimkarn S, Dumic M, Obeid J, Azar MR, Najmabadi H, et al. Ethnic-specific distribution of mutations in 716 patients with congenital adrenal hyperplasia owing to 21-hydroxylase deficiency. Mol Genet Metab. 2007;90(4):414–21.
5. Merke DP, Auchus RJ. Congenital adrenal hyperplasia due to 21-hydroxylase deficiency. N Engl J Med. 2020;383(13):1248–61.
6. Higashi Y, Yoshioka H, Yamane M, Gotoh O, Fujii-Kuriyama Y. Complete nucleotide sequence of two steroid 21-hydroxylase genes tandemly arranged in human chromosome: a pseudogene and a genuine gene. Proc Natl Acad Sci USA. 1986;83(9):2841–5.
7. Yang Z, Mendoza AR, Welch TR, Zipf WB, Yu CY. Modular variations of the human major histocompatibility complex class III genes for serine/threonine kinase RP, complement component C4, steroid 21-hydroxylase CYP21, and tenascin TNX (the RCCX module). A mechanism for gene deletions and disease associations. J Biol Chem. 1999;274(17):12147–56.
8. Blanchong CA, Zhou B, Rupert KL, Chung EK, Jones KN, Sotos JF, et al. Deficiencies of human complement component C4A and C4B and

heterozygosity in length variants of RP-C4-CYP21-TNX (RCCX) modules in Caucasians. The load of RCCX genetic diversity on major histocompatibility complex-associated disease. J Exp Med. 2000;191(12):2183–96.
9. Carrozza C, Foca L, De Paolis E, Concolino P. Genes and pseudogenes: complexity of the RCCX locus and disease. Front Endocrinol. 2021;12:709758.
10. Doleschall M, Luczay A, Koncz K, Hadzsiev K, Erhardt É, Szilágyi Á, et al. A unique haplotype of RCCX copy number variation: from the clinics of congenital adrenal hyperplasia to evolutionary genetics. Eur J Hum Genet EJHG. 2017;25(6):702–10.
11. Concolino P, Costella A. Congenital adrenal hyperplasia (CAH) due to 21-hydroxylase deficiency: a comprehensive focus on 233 pathogenic variants of CYP21A2 gene. Mol Diagn Ther. 2018;22(3):261–80.
12. Finkielstain GP, Chen W, Mehta SP, Fujimura FK, Hanna RM, Van Ryzin C, et al. Comprehensive genetic analysis of 182 unrelated families with congenital adrenal hyperplasia due to 21-hydroxylase deficiency. J Clin Endocrinol Metab. 2011;96(1):E161-172.
13. Pignatelli D, Carvalho BL, Palmeiro A, Barros A, Guerreiro SG, Macut D. The complexities in genotyping of congenital adrenal hyperplasia: 21-hydroxylase deficiency. Front Endocrinol. 2019;10:432.
14. Chen W, Xu Z, Sullivan A, Finkielstain GP, Van Ryzin C, Merke DP, et al. Junction site analysis of chimeric CYP21A1P/CYP21A2 genes in 21-hydroxylase deficiency. Clin Chem. 2012;58(2):421–30.
15. Merke DP, Chen W, Morissette R, Xu Z, Van Ryzin C, Sachdev V, et al. Tenascin-X haploinsufficiency associated with Ehlers–Danlos syndrome in patients with congenital adrenal hyperplasia. J Clin Endocrinol Metab. 2013;98(2):E379-387.
16. Paragliola RM, Perrucci A, Foca L, Urbani A, Concolino P. Prevalence of CAH-X syndrome in Italian patients with congenital adrenal hyperplasia (CAH) due to 21-hydroxylase deficiency. J Clin Med. 2022;11(13):3818.
17. Chen W, Kim MS, Shanbhag S, Arai A, VanRyzin C, McDonnell NB, et al. The phenotypic spectrum of contiguous deletion of CYP21A2 and tenascin XB: quadricuspid aortic valve and other midline defects. Am J Med Genet Part A. 2009;149a(12):2803–8.
18. Ezquieta B, Beneyto M, Muñoz-Pacheco R, Barrio R, Oyarzabal M, Lechuga JL, et al. Gene duplications in 21-hydroxylase deficiency: the importance of accurate molecular diagnosis in carrier detection and prenatal diagnosis. Prenat Diagn. 2006;26(12):1172–8.
19. Kleinle S, Lang R, Fischer GF, Vierhapper H, Waldhauser F, Födinger M, et al. Duplications of the functional CYP21A2 gene are primarily restricted to Q318X alleles: evidence for a founder effect. J Clin Endocrinol Metab. 2009;94(10):3954–8.
20. White PC, Curnow KM, Pascoe L. Disorders of steroid 11 beta-hydroxylase isozymes. Endocr Rev. 1994;15(4):421–38.
21. Xie H, Yin H, Ye X, Liu Y, Liu N, Zhang Y, et al. Detection of small CYP11B1 deletions and one founder chimeric CYP11B2/CYP11B1 gene in 11β-hydroxylase deficiency. Front Endocrinol. 2022;13:882863.
22. Xu Z, Chen W, Merke DP, McDonnell NB. Comprehensive mutation analysis of the CYP21A2 gene: an efficient multistep approach to the molecular diagnosis of congenital adrenal hyperplasia. J Mol Diagn JMD. 2013;15(6):745–53.
23. Gao Y, Lu L, Yu B, Mao J, Wang X, Nie M, et al. The prevalence of the chimeric TNXA/TNXB gene and clinical symptoms of Ehlers–Danlos syndrome with 21-hydroxylase deficiency. J Clin Endocrinol Metabol. 2020;105(7):2288–99.
24. Ewans LJ, Schofield D, Shrestha R, Zhu Y, Gayevskiy V, Ying K, et al. Whole-exome sequencing reanalysis at 12 months boosts diagnosis and is cost-effective when applied early in Mendelian disorders. Genet Med Off J Am Coll Med Genet. 2018;20(12):1564–74.
25. Ewans LJ, Minoche AE, Schofield D, Shrestha R, Puttick C, Zhu Y, et al. Whole exome and genome sequencing in mendelian disorders: a diagnostic and health economic analysis. Eur J Hum Genet EJHG. 2022;30(10):1121–31.
26. Hiz Kurul S, Oktay Y, Töpf A, Szabó NZ, Güngör S, Yaramis A, et al. High diagnostic rate of trio exome sequencing in consanguineous families with neurogenetic diseases. Brain J Neurol. 2022;145(4):1507–18.
27. Gangodkar P, Khadilkar V, Raghupathy P, Kumar R, Dayal AA, Dayal D, et al. Clinical application of a novel next generation sequencing assay for CYP21A2 gene in 310 cases of 21- hydroxylase congenital adrenal hyperplasia from India. Endocrine. 2021;71(1):189–98.
28. Turan I, Tastan M, Boga DD, Gurbuz F, Kotan LD, Tuli A, et al. 21-Hydroxylase deficiency: mutational spectrum and genotype–phenotype relations analyses by next-generation sequencing and multiplex ligation-dependent probe amplification. Eur J Med Genet. 2020;63(4):103782.
29. Wang W, Han R, Yang Z, Zheng S, Li H, Wan Z, et al. Targeted gene panel sequencing for molecular diagnosis of congenital adrenal hyperplasia. J Steroid Biochem Mol Biol. 2021;211:105899.
30. Wenger AM, Peluso P, Rowell WJ, Chang PC, Hall RJ, Concepcion GT, et al. Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. Nat Biotechnol. 2019;37(10):1155–62.
31. Mastrorosa FK, Miller DE, Eichler EE. Applications of long-read sequencing to Mendelian genetics. Genome Med. 2023;15(1):42.
32. Liu Y, Chen M, Liu J, Mao A, Teng Y, Yan H, et al. Comprehensive analysis of congenital adrenal hyperplasia using long-read sequencing. Clin Chem. 2022;68(7):927–39.
33. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics (Oxford, England). 2009;25(14):1754–60.
34. Schouten JP, McElgunn CJ, Waaijer R, Zwijnenburg D, Diepvens F, Pals G. Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. Nucleic Acids Res. 2002;30(12):e57.
35. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med Off J Am Coll Med Genet. 2015;17(5):405–24.
36. Richards CS, Bale S, Bellissimo DB, Das S, Grody WW, Hegde MR, et al. ACMG recommendations for standards for interpretation and reporting of sequence variations: revisions 2007. Genet Med Off J Am Coll Med Genet. 2008;10(4):294–300.
37. Plon SE, Eccles DM, Easton D, Foulkes WD, Genuardi M, Greenblatt MS, et al. Sequence variant classification and reporting: recommendations for improving the interpretation of cancer susceptibility genetic test results. Hum Mutat. 2008;29(11):1282–91.
38. Miller DE, Sulovari A, Wang T, Loucks H, Hoekzema K, Munson KM, et al. Targeted long-read sequencing identifies missing disease-causing variation. Am J Hum Genet. 2021;108(8):1436–49.
39. Liu L, Li Y, Li S, Hu N, He Y, Pong R, et al. Comparison of next-generation sequencing systems. J Biomed Biotechnol. 2012;2012:251364.
40. Cheng C, Fei Z, Xiao P. Methods to improve the accuracy of next-generation sequencing. Front Bioeng Biotechnol. 2023;11:982111.
41. Araujo RS, Billerbeck AE, Madureira G, Mendonca BB, Bachega TA. Substitutions in the CYP21A2 promoter explain the simple-virilizing form of 21-hydroxylase deficiency in patients harbouring a P30L mutation. Clin Endocrinol. 2005;62(2):132–6.
42. Baumgartner-Parzer S, Witsch-Baumgartner M, Hoeppner W. EMQN best practice guidelines for molecular genetic testing and reporting of 21-hydroxylase deficiency. Eur J Hum Genet EJHG. 2020;28(10):1341–67.
43. Fanis P, Skordis N, Toumba M, Picolos M, Tanteles GA, Neocleous V, et al. The pathogenic p.Gln319Ter variant is not causing congenital adrenal hyperplasia when inherited in one of the duplicated CYP21A2 genes. Front Endocrinol. 2023;14:1156616.
44. Li Z, Liang Y, Du C, Yu X, Hou L, Wu W, et al. Clinical applications of genetic analysis and liquid chromatography tandem-mass spectrometry in rare types of congenital adrenal hyperplasia. BMC Endocr Disord. 2021;21(1):237.
45. Liang Q, Gu W, Chen P, Li Y, Liu Y, Tian M, et al. A more universal approach to comprehensive analysis of thalassemia alleles (CATSA). J Mol Diagn JMD. 2021;23(9):1195–204.
46. Xu L, Mao A, Liu H, Gui B, Choy KW, Huang H, et al. Long-molecule sequencing: a new approach for identification of clinically significant DNA variants in α-thalassemia and β-thalassemia carriers. J Mol Diagn JMD. 2020;22(8):1087–95.
47. Nomura F, Shimizu A, Togi S, Ura H, Niida Y. SNP array screening and long range PCR-based targeted next generation sequencing for autosomal recessive disease with consanguinity: insight from a case of xeroderma pigmentosum group C. Genes. 2023;14(11):2079.
48. Ai X, Li B, Xu Z, Liu J, Qin T, Li Q, et al. Multiplex ligation-dependent probe amplification and fluorescence in situ hybridization for detecting

chromosome abnormalities in myelodysplastic syndromes: a retrospective study. Medicine. 2021;100(18):e25768.
49. Li H, Zhu X, Yang Y, Wang W, Mao A, Li J, et al. Long-read sequencing: an effective method for genetic analysis of CYP21A2 variation in congenital adrenal hyperplasia. Clin Chim Acta Int J Clin Chem. 2023;547:117419.

Publisher's Note